

Aplicación web interactiva en "Shiny"
para el análisis espacio-temporal de riesgos de
mortalidad por cáncer



Grado en Ingeniería Informática

Trabajo Fin de Grado

Alumno: Javier Miguel Carrillo Corpas
Tutores : Aritz Adin Urtasun y
María Dolores Ugarte Martínez
Pamplona, 27 de enero de 2017

Agradecimientos

Quiero agradecer a mis padres la oportunidad que me han brindado para realizar los estudios de Grado en Ingeniería Informática. A todos los compañeros y profesores por su ayuda durante mi paso por la universidad y a mis tutores por la ayuda que me han prestado para la realización de este trabajo.

Resumen

El objetivo de este trabajo es la implementación de una aplicación web utilizando "Shiny". "Shiny" es un entorno de trabajo para el desarrollo de aplicaciones web en lenguaje R. La aplicación permite un acceso sencillo a una serie de herramientas para la representación cartográfica de enfermedades. Normalmente se representan los riesgos o tasas suavizadas en lugar de los datos brutos. Estas tasas se suavizan usando modelos estadísticos cada vez más complejos, debido a la variabilidad que muestran al analizar enfermedades raras o cuando se calculan sobre áreas pequeñas. La aplicación da acceso a la implementación de distintos modelos estadísticos, desarrollados por el grupo Estadística Espacial de la UPNA. Estos modelos utilizan la metodología INLA (Integrated Nested Laplace Approximations) para suavizar los riesgos o tasas.

Palabras clave

Aplicación web, INLA, modelos espacio-temporales, R, Shiny.

ÍNDICE

AGRADECIMIENTOS.....	3
RESUMEN	5
CAPÍTULO 1. INTRODUCCIÓN	9
1.1 Visión general.....	9
1.2 Objetivos	9
CAPÍTULO 2. CONTEXTO TECNOLÓGICO	11
2.1 Aplicaciones web.....	11
2.2 Lenguajes propios de internet.....	12
2.3 Lenguaje R.....	12
2.4 IDE RStudio.....	12
2.5 Shiny.....	13
2.6 Bootstrap.....	13
2.7 Shinythemes.....	14
2.8 ShinyBS.....	14
2.9 ShinySky	14
2.10 INLA.....	14
CAPÍTULO 3. LA APLICACIÓN	15
3.1 Metodología.....	15
3.2 Diseño	15
3.3 Implementación	15
3.4 Resultados.....	17
3.5 Lineas futuras	17
3.6 Conclusiones	18

CAPÍTULO 4. MANUAL DE USUARIO	19
4.1 Introducción	21
4.2 Entrada de datos	23
4.3 Selección de variables	25
4.4 Análisis descriptivo	29
4.5 Selección del modelo.....	33
4.6 Ejemplo de uso	37
ANEXOS.....	49
BIBLIOGRAFÍA.....	51

CAPÍTULO 1. INTRODUCCIÓN

1.1 Visión general

La representación cartográfica de enfermedades, también conocida como epidemiología espacial, estudia los patrones geográficos de mortalidad o incidencia de enfermedades y su evolución en el tiempo, permitiendo así la detección de regiones con riesgos extremos. Los modelos estadísticos utilizados son complejos y por ello, resultan de difícil aplicación por los profesionales de la salud pública en general. Se espera que la aplicación desarrollada facilite el uso de los modelos y sea una herramienta útil para la representación cartográfica de enfermedades.

Puesto que estos modelos [17] se implementaron usando el lenguaje de programación R, se decidió usar "Shiny" un entorno de trabajo para desarrollar aplicaciones web programadas en R, que permite la construcción de aplicaciones web sin la necesidad de usar HTML, CSS o JavaScript. Aunque estos lenguajes se pueden usar para personalizar las aplicaciones o añadir funcionalidades extra.

1.2 Objetivos

El objetivo es desarrollar una aplicación web que proporcione a los epidemiólogos y a los investigadores en salud pública en general, una herramienta de uso amigable que permita estimar riesgos o tasas de mortalidad (o de incidencia de enfermedades). La herramienta desarrollada permitirá a su vez la representación gráfica de los resultados en forma de mapas, de modo que sea posible visualizar el patrón espacial, temporal y espacio-temporal de la enfermedad objeto de estudio.

Este trabajo de fin de grado se centra en la implementación inicial de la aplicación web, permitiendo el acceso a los modelos y definiendo el esquema de la misma.

En el capítulo 2 se explica el contexto tecnológico de la aplicación. La aplicación en sí se describe en el capítulo 3. El capítulo 4 contiene un manual de usuario con un ejemplo de uso de la aplicación. Se incluye además un anexo que recoge una copia del certificado de presentación de esta aplicación en las Jornadas de Usuarios de R que se celebraron en Albacete el pasado noviembre de 2016. Por último se proporciona la bibliografía utilizada en la realización de este trabajo.

CAPÍTULO 2. CONTEXTO TECNOLÓGICO

En este capítulo se hablará de las tecnologías utilizadas en el desarrollo de la aplicación, analizando brevemente sus ventajas.

2.1 Aplicaciones web

Una aplicación web es una aplicación software accesible a través de internet mediante un navegador web. Estas aplicaciones son muy populares hoy en día para los usuarios no expertos, debido a la facilidad de su uso, ya que no es preciso instalar nada en el ordenador, simplemente acceder a la aplicación a través del navegador. Para el cliente también es una ventaja el bajo consumo de recursos, ya que la mayor parte del tiempo estos se consumen en el servidor, que generalmente tiene mucha más potencia de cómputo que cualquier ordenador personal. Además se puede acceder desde cualquier dispositivo con conexión a internet, ya sea un ordenador, un smartphone o una tablet. Esta es una de las mayores ventajas para los proveedores de aplicaciones web, la independencia del sistema operativo del usuario. Otra ventaja de las aplicaciones web es que son más fáciles de mantener, ya que las actualizaciones sólo deben instalarse en el servidor. Como desventajas cabe destacar que las aplicaciones web no ofrecen tantas funcionalidades como una aplicación de escritorio, ya que se ejecutan desde un navegador y no desde el sistema operativo. Además se necesita que los dispositivos cliente y servidor estén conectados, lo que hace que la disponibilidad de la aplicación dependa del proveedor de conexión a internet y ni usuario ni proveedor pueden hacer nada para solucionar un problema en este nivel. Puesto que uno de los objetivos de la aplicación era permitir el acceso a los modelos de forma sencilla se eligió una aplicación web sobre una de escritorio.

2.2 Lenguajes propios de internet

En las páginas y aplicaciones web se ha estandarizado el uso de ciertos lenguajes, tales como HTML, CSS, PHP o Javascript entre otros. En el desarrollo de esta aplicación no se han usado de forma extensa, ya que el entorno de trabajo "Shiny" [9] no requiere ni conocimiento ni uso de estos lenguajes, si bien permite su uso para personalizar las aplicaciones o implementar alguna funcionalidad específica. "Shiny" permite traducir el código en lenguaje R a una aplicación que usa HTML, CSS y Javascript para funcionar de forma sencilla.

2.3 Lenguaje R

R [20] es un entorno y lenguaje de programación diseñado por Ross Ihaka y Robert Gentleman. R es una implementación libre de lenguaje S, pero con un enfoque al análisis estadístico. Actualmente R es uno de los lenguajes más utilizados en investigación por los estadísticos. Al tratarse de un proyecto abierto y colaborativo, dispone de una gran cantidad de bibliotecas o paquetes, que los usuarios pueden publicar, con funcionalidades extra como puede ser la graficación. El repositorio oficial de paquetes [15] cuenta actualmente con más de 9000 paquetes. R es parte del sistema GNU, pero también está disponible para sistemas operativos Windows o Macintosh. La última versión estable de R es la 3.3.2 y fue publicada el 31/10/2016.

2.4 IDE RStudio

Para facilitar el trabajo con el lenguaje R se usó el entorno de desarrollo integrado (IDE) RStudio [22]. RStudio está disponible en código abierto bajo licencia AGPL v3 y en edición comercial y funciona en sistemas operativos Windows, Macintosh y Linux. Para el desarrollo de esta aplicación las herramientas más usadas de RStudio fueron el editor de sintaxis y el debugger.

2.5 Shiny

Shiny es, como se ha expuesto anteriormente, un entorno de trabajo para construir aplicaciones web en R de forma sencilla, sin necesitar conocimientos de HTML, CSS o Javascript. Shiny es un paquete de R creado en 2012 por RStudio, por lo que su instalación es tan sencilla como la de cualquier otro paquete de R, mediante la instrucción `install.packages("shiny")`. Entre sus características se encuentran la programación reactiva, HTML5, CSS3 y Javascript. Shiny se encuentra en su versión 0.14, publicada el 12/09/2016.

2.6 Bootstrap

Por defecto Shiny utiliza una plantilla básica de Twitter Bootstrap [18] para crear la interfaz de usuario, pero podemos descargar otras plantillas o crear una propia para personalizar nuestra aplicación.

Twitter Bootstrap es un entorno de trabajo desarrollado por empleados de Twitter para fomentar la consistencia entre las herramientas internas, de forma que todas siguieran el mismo estilo. En 2011 Twitter liberó Bootstrap como código abierto, permitiendo que cualquiera lo usara para diseñar sus sitios o aplicaciones web. Contiene elementos de diseño basado en HTML, CSS y Javascript. Una de las mayores ventajas de Bootstrap es que permite crear interfaces web con CSS y JavaScript que adaptan la interfaz dependiendo del tamaño del dispositivo en el que se visualice de forma nativa, es decir, automáticamente se adapta al tamaño de un ordenador o de una tablet sin que el usuario tenga que hacer nada. Esto se denomina diseño adaptable o Responsive Design.

Actualmente es el proyecto más popular en GitHub y se encuentra en la versión 3.3.6, que cuenta con un soporte incompleto para HTML5 y CSS, pero es compatible con la mayoría de los navegadores web.

2.7 Shinythemes

En lugar de usar un CSS propio para la interfaz de la aplicación se ha usado el paquete de R shinythemes [8], publicado por los creadores del propio Shiny. Shinythemes ofrece una serie de estilos básicos para aplicaciones Shiny. Shinythemes es compatible con CSS. En el desarrollo de esta aplicación se ha usado CSS para resaltar los avisos al usuario. La última actualización de Shinythemes (1.1.1) está fechada el 13/10/2016.

2.8 ShinyBS

ShinyBS [3] es una librería que agrega componentes de Twitter Bootstrap a una aplicación Shiny para añadir funcionalidades extra. En concreto se utilizó para estructurar la sección de ayuda de la aplicación. Se encuentra actualmente en la versión 0.61 publicada el 31/03/2015 y está disponible en CRAN, el repositorio oficial de paquetes de R.

2.9 ShinySky

ShinySky [2] es una librería que agrega componentes de interfaz de usuario a una aplicación Shiny. En concreto se utilizó para añadir un mensaje de espera mientras se calcula el modelo. Se encuentra disponible en GitHub.

2.10 INLA

R-INLA [19] es un paquete de R que permite el uso de la aproximación integrada de Laplace anidada [10] [11] [12] [16] para estimar modelos espaciales y espacio-temporales

Debido a la alta carga computacional de INLA la estimación de los modelos se ejecutará de forma remota en un servidor proporcionado por la UPNA. Para la ejecución remota de INLA es necesaria la comunicación segura entre las dos máquinas, para ello R-INLA utiliza SSH, un protocolo de comunicación segura que permite la ejecución de órdenes y comandos de consola.

CAPÍTULO 3. LA APLICACIÓN

En este capítulo se explica el proceso de implementación de la aplicación así como los resultados obtenidos y las líneas futuras de desarrollo.

3.1 Metodología

Para la realización de este proyecto se ha usado una metodología ágil. Las metodologías ágiles enfatizan la comunicación cara a cara sobre la documentación y enfatizan el software funcional como primera medida del progreso.

El proceso consistió en una exposición general de la aplicación. Un periodo de aprendizaje para familiarizarse con Shiny, realizando tutoriales on-line y buscando en la galería oficial elementos que pudieran ser de utilidad para la aplicación. Una vez iniciado el desarrollo de la aplicación se realizaron reuniones, más o menos regulares según la disponibilidad de los implicados, en las que se exponía el trabajo realizado desde la última reunión. En estas reuniones se discutían también los objetivos para la próxima reunión, corrección de errores detectados y cambios de diseño.

3.2 Diseño

El diseño de la aplicación se basa en la estructura básica de las aplicaciones "Shiny". Esta estructura separa el código de la aplicación en dos archivos según su función. El archivo `ui.R` que recoge la código pertinente a la interfaz de usuario y el archivo `server.R` que contiene las funcionalidades de la aplicación.

3.3 Implementación

Tras la realización de los tutoriales en Shiny, se comenzó a construir la aplicación.

Primero se decidió la interfaz de usuario. Para ello se realizaron una serie de pruebas, usando las diferentes posibilidades que ofrece "Shiny" para la creación de interfaces de usuario. Estas pruebas se expusieron y se eligió la opción que se consideró que separaba mejor las diferentes partes de la aplicación. El diseño elegido separa la aplicación en pestañas accesibles desde un menú situado en la parte superior. Además cada pestaña se separó en un menú lateral y una sección central.

Una vez definida la interfaz de usuario se pasó a implementar las funcionalidades necesarias para un primer ejemplo de la aplicación. Estas funcionalidades son: por un lado permitir la carga de los datos del usuario, así como la cartografía asociada, y por otro permitir al usuario interactuar con los datos cargados, seleccionando las variables de interés. En particular, la aplicación permite calcular y representar las tasas brutas, la evolución temporal de las tasas brutas y las razones de mortalidad estandarizadas (SMR). Para el cálculo y representación de los SMR se utilizan casillas de verificación (checkbox) para que el usuario seleccione los años que quiere representar. Posteriormente se añadió la funcionalidad de poder descargar un mapa de los SMRs en formato ".pdf". Durante esta fase se utilizó código en R, así como un conjunto de datos reales de mortalidad por cáncer, facilitados por los tutores para la realización de pruebas y para comprobar el correcto funcionamiento de la aplicación.

La siguiente fase se centró en agregar la funcionalidad de calcular los casos esperados. Este cálculo depende de una serie de variables auxiliares que el usuario debe seleccionar. Se decidió que la aplicación permitiese calcular los casos esperados o que dichos casos fueran facilitados por el usuario. El código del cálculo de los casos esperados fue facilitado por los tutores en forma de función de R que la aplicación debía usar.

La última fase se centró en agregar la funcionalidad de la estimación de los modelos usando INLA. Durante el transcurso de esta fase también se llevó a cabo una reestructuración del código de la aplicación para hacerlo mas legible e intuitivo. También se agregaron componentes visuales para embellecer la aplicación y se agregaron los componentes de ayuda y manual.

3.4 Resultados

El resultado final de la aplicación permite al usuario crear y descargar varios mapas espacio-temporales y visualizar el modelo obtenido.

La versión entregada de la aplicación contiene:

- Los ficheros ui.R y server.R.
Contienen la lógica de la aplicación y de la interfaz de usuario.
- El código en R proporcionados al alumno.
Contienen la implementación de las funciones desarrolladas por los tutores.
- Los datos de prueba proporcionados al alumno.
Contienen los datos y cartografía usados por el alumno para probar el funcionamiento de la aplicación.
- Las subcarpetas "pdfs", "temp", "downloads" y "www"
guardan los mapas creados, los datos temporales de la ejecución de INLA, el manual y los datos de ejemplo, y el CSS propio respectivamente.

3.5 Líneas futuras

La aplicación, si bien tiene el esqueleto terminado, aún presenta ciertas características sin probar y ciertas funcionalidades que podrían agregarse. Algunas opciones para la estimación de los modelos no se han probado. Estas opciones se encuentran deshabilitadas. Es necesario también una comprobación más exhaustiva de la aplicación para que funcione correctamente en situaciones prácticas.

El despliegue final puede hacerse de dos formas. La primera opción consiste en descargar e instalar Shiny server en un ordenador propio. La segunda opción consiste en registrarse en shinyapps.io y subir la aplicación a la nube, de forma que la aplicación se ejecuta en un servidor externo. Se recomienda la primera opción por posibles incompatibilidades al usar SSH en la parte de generación del modelo con INLA.

Como mejoras posibles se podría abrir un canal de feedback propio de la aplicación, por ejemplo una cuenta de gmail y automatizar en el servidor final de la aplicación la destrucción de los mapas y datos temporales generados por la aplicación.

3.6 Conclusiones

Se ha aprendido a usar un IDE (Entorno de Desarrollo Integrado) nuevo y se han repasado los conocimientos de R obtenidos en el grado. Se ha aprendido a usar Shiny para la creación de aplicaciones web. Se ha obtenido experiencia en el desarrollo de aplicaciones en solitario, sin un profesor que ya sabe el resultado final y que puede guiarnos, debiendo el alumno buscar los recursos y soluciones a ciertos problemas planteados por el cliente por su cuenta.

Como apunte final los tutores de este trabajo presentaron una demo de esta aplicación en las VIII Jornadas de Usuarios de R en Albacete.

CAPÍTULO 4. MANUAL DE USUARIO

Aplicación Shiny para la representación espacio-temporal de casos de enfermedades

Manual de usuario



Autor :

Javier Miguel Carrillo Corpas

Supervisores:

Aritz Adin Urtasun

María Dolores Ugarte Martínez

Capítulo 4.1. Introducción

En este documento se explican los pasos a seguir para utilizar correctamente la aplicación web.

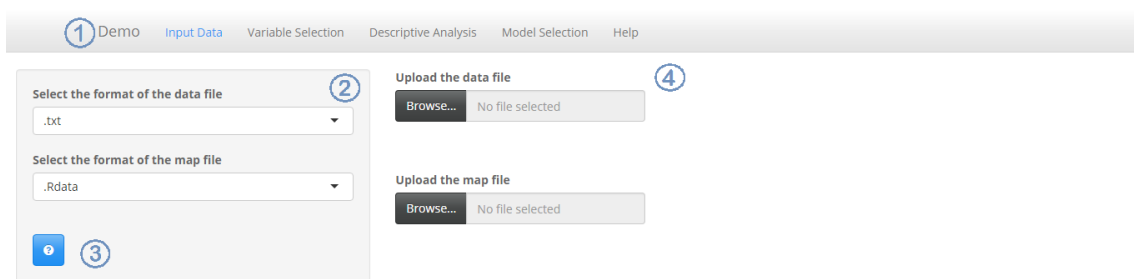
El objetivo de la aplicación es facilitar la representación cartográfica de enfermedades, así como poner a disposición del público los modelos estadísticos desarrollados por el grupo de Estadística Espacial de la UPNA.

La aplicación permite al usuario introducir tanto datos espaciales como espacio-temporales (datos agregados por áreas e instantes de tiempo) de mortalidad o incidencia, así como la cartografía asociada a los mismos. Una vez seleccionadas las variables de interés, la aplicación permite realizar un análisis descriptivo de los datos. Concretamente, permite representar la evolución temporal y la distribución geográfica de las tasas brutas o las razones de mortalidad estandarizadas (SMR). Finalmente, la aplicación permite obtener estimaciones de los riesgos relativos de mortalidad o incidencia ajustando distintos modelos de suavizado a través del paquete R-INLA. Todos los resultados y gráficos generados por la aplicación pueden ser exportados por el usuario.

Capítulo 4.2. Entrada de datos

La aplicación permite al usuario importar conjuntos de datos en los siguientes formatos: archivos de texto ".txt", de valores separados por comas, ".csv" o archivos de datos binarios de R ".Rdata". La cartografía asociada a los ficheros de entrada puede facilitarse en formato .rds o en .Rdata. La aplicación sólo admite la carga de archivos con un tamaño menor a 30MB.

La vista inicial de la aplicación es la siguiente:



En ella podemos:

En la parte superior:

Navegar por la aplicación con el menú (1).

En la parte izquierda:

Seleccionar el formato de los archivos (2).

Mostrar la ayuda (3).

En la parte central:

Subir nuestros archivos (4).

Capítulo 4.3. Selección de Variables

Si no se ha cargado alguno de los ficheros antes de pasar a este paso aparecerá un mensaje de error en la parte central avisando de qué archivo falta:

The screenshot shows the 'Variable Selection' step of a software interface. The sidebar on the left contains the following options:

- Select the time variable: [dropdown menu]
- Select the area variable: [dropdown menu]
- Select the total population variable: [dropdown menu]
- Select the cases variable: [dropdown menu]
- Select the area variable in the map: [dropdown menu]
- Select the expected cases variable: [dropdown menu]
- ☐ Calculate the expected cases with our models usign auxiliary variables

At the bottom of the sidebar is a blue button with a white 'e' icon. The main area on the right displays a red error message: "Missing both Files".

Si los ficheros se han subido correctamente se verá lo siguiente:

The screenshot shows the 'Variable Selection' step of a software interface with successful file uploads. The sidebar on the left contains the following options:

- Select the time variable: [dropdown menu with value 'ano']
- Select the area variable: [dropdown menu with value 'prov']
- Select the total population variable: [dropdown menu with value 'pob']
- Select the cases variable: [dropdown menu with value 'pob']
- Select the area variable in the map: [dropdown menu with value 'NAME']
- Select the expected cases variable: [dropdown menu with value 'esp']
- ☐ Calculate the expected cases with our models usign auxiliary variables

At the bottom of the sidebar is a blue button with a white 'e' icon. The main area on the right displays the following information:

- WARNING Same variable for several fields** (3)
- YEAR -> Values: 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
- AREA -> Values: 1 2 3 4 5 6 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 36 37 39 40 41 42 43 44 45 46 47 48 49 50
- POPULATION -> Range: (2)
1496 286088
- CASES -> Range:
1496 286088
- WARNING the area identifiers do not match** (3)

En esta sección podemos:

En la parte superior:

Navegar por la aplicación con el menú.

En la parte izquierda:

Seleccionar las variables a usar (1).

Mostrar la ayuda.

En la parte central:

Visualizar un resumen de lo seleccionado (2) y un aviso si seleccionas la misma variable dos veces o si los identificadores del mapa no se corresponden a los de los datos introducidos (3).

En este paso debemos seleccionar las variables a usar en el menú de la izquierda. Los nombres de estas variables se corresponden a las columnas de los ficheros cargados. Deberemos seleccionar obligatoriamente las variables correspondientes a tiempo, área, población, casos y el identificador del área en el mapa, que debe coincidir con el identificador del área de los datos.

Después, si los casos están recogidos en una variable del fichero de datos, seleccionamos dicha variable y continuamos haciendo clic sobre "Descriptive Analysis".

Si por el contrario queremos que la aplicación calcule, mediante el método de estandarización indirecta, los casos esperados en función de ciertas variables auxiliares, por ejemplo grupos de edad o sexo, marcamos la opción de "calcular los casos esperados con nuestros modelos" (1). Veremos cómo desaparece el desplegable de selección de los casos esperados y aparecen dos botones nuevos. Agregamos tantas variables auxiliares como necesitemos mediante el botón "Agregar variables auxiliar" (2) y seleccionamos las variables en los desplegables que aparecerán (3).

Select the area variable in the map

ID.area

1

☒ Calculate the expected cases with our models using auxiliary variables

Add new auxiliary variable

Delete auxiliary variable

2

Variable number 1

edadgr

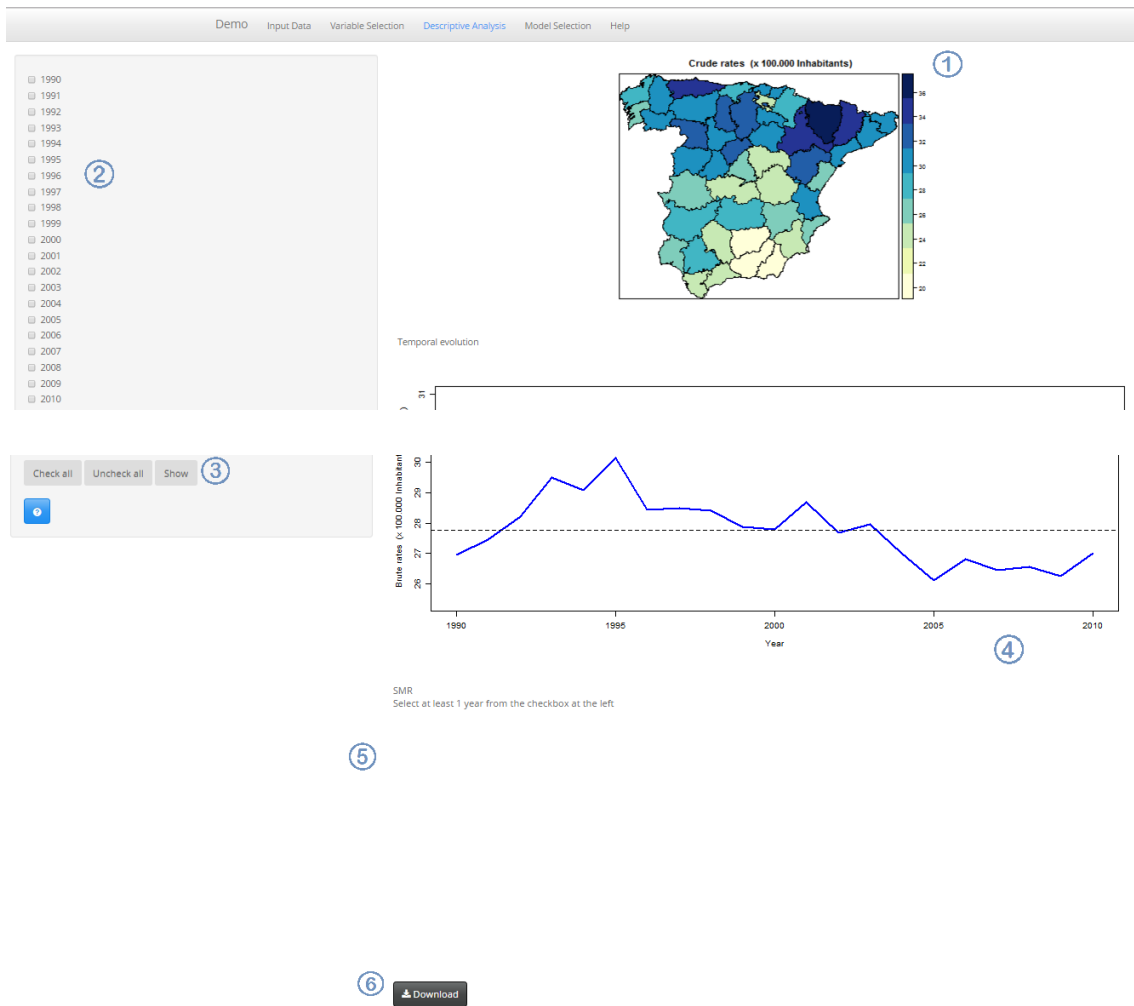
3

?

Y continuamos haciendo clic sobre "Descriptive Analisis" en el menú superior.

Capítulo 4.4. Análisis descriptivo

Al llegar a este punto, si todo ha funcionado correctamente, tras unos instantes veremos aparecer lo siguiente:



En esta sección podemos:

En la parte superior:

Navegar por la aplicación con el menú.

En la parte izquierda:

Seleccionar los años de los cuales queremos representar los riesgos de mortalidad estandarizados (SMR) (1).

Indicar a la aplicación que ya hemos seleccionado los años que queremos, botón "show" (3), y mostrarlos en pantalla en el espacio indicado (5).

Mostrar la ayuda.

En la parte central:

Visualizar la representación geográfica de las tasas brutas (1).

Visualizar una representación temporal (4).

Visualizar ,una vez seleccionados los años, los SMR (5).

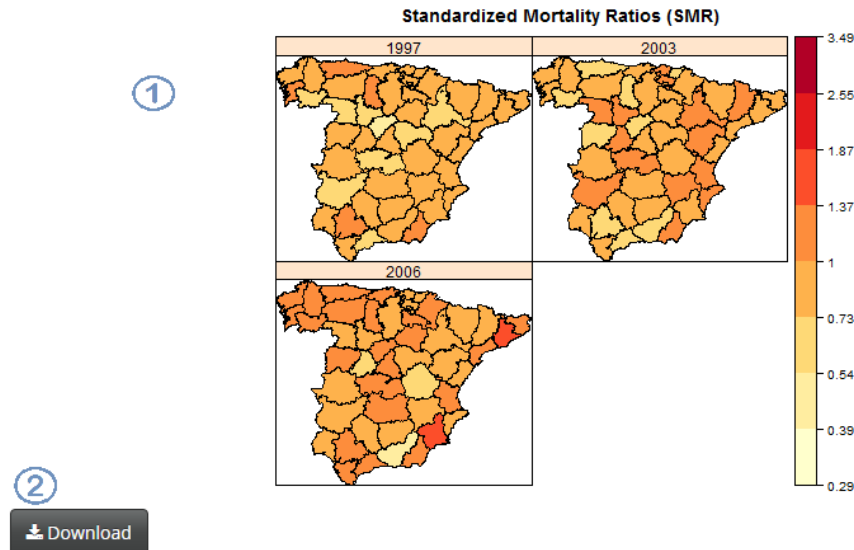
Descargar el gráfico con los SMR representados en formato ".pdf" (6).

En esta sección las representaciones geográfica y temporal se muestran de forma automática, es decir, no tenemos que indicarle nada más a la aplicación.

Debemos seleccionar los años, o unidades de tiempo, que nos interesan estudiar. Podemos usar los botones "check all" y "uncheck all" para facilitar la selección.

Una vez seleccionados los años hacemos clic en "show" y esperamos a que aparezcan los SMR de esta forma (1):

SMR

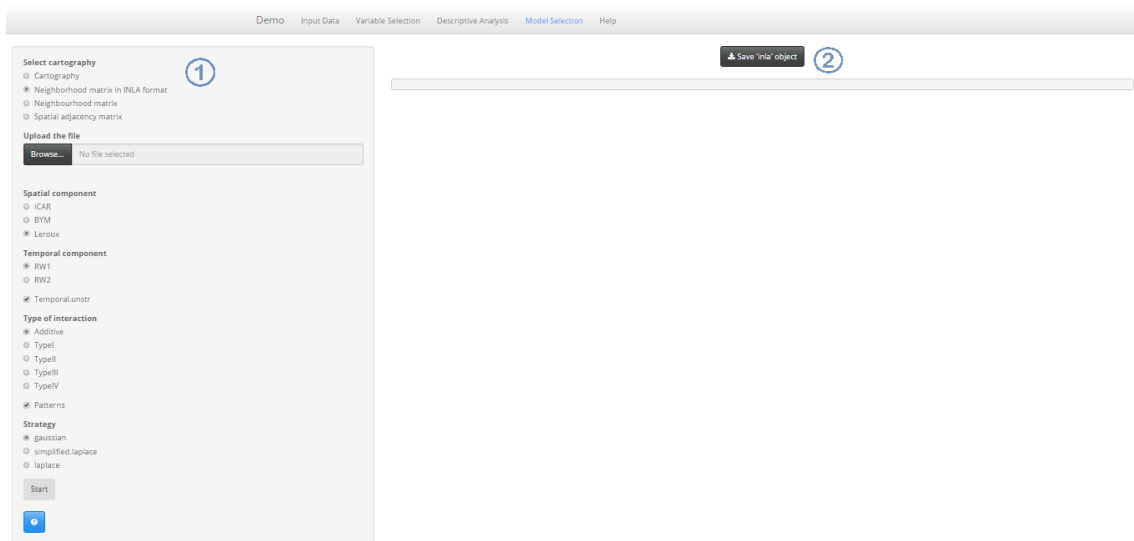


Haciendo clic sobre el botón "Download" (2) podremos descargar la figura (1) en formato .pdf. Si se desea descargarla en otro formato es posible hacerlo con un clic (derecho) sobre la figura y seleccionando la opción "guardar como".

Podemos continuar haciendo clic sobre "Model Selection" en el menú superior.

Capítulo 4.5. Selección del modelo

Al entrar en esta sección veremos lo siguiente:



En esta sección podemos:

En la parte superior:

Navegar por la aplicación con el menú.

En la parte izquierda:

Seleccionar las opciones para el cálculo del modelo (1).

Mostrar la ayuda.

En la parte central:

Visualizar y descargar el modelo calculado (2).

En esta sección debemos seleccionar los parámetros que deseamos que la aplicación use para ajustar los modelos.

Primero debemos seleccionar si queremos usar la cartografía ya cargada. A continuación debemos cargar la matriz que especifica las vecindades. Hay tres posibilidades distintas:

- 1.- Cargar la matriz en formato INLA, archivo de extensión ".inla", una matriz que en su primera fila indica el número de áreas y una fila por cada área en la que se indica con que otras áreas delimita.
- 2.- Cargar la matriz de adyacencia, una matriz de $N \times N$ en la que N es el número de áreas y cada valor M_{ij} es 1 o 0 dependiendo de si las áreas i, j se delimitan o no. Esta matriz es un objeto de clase matrix de R y su extensión es ".txt".
- 3.- Cargar la matriz de vecindades, una matriz de $N \times N$ en la que N es el número de áreas y en la que cada valor de la diagonal M_{ii} es el número de áreas con las que delimita y el resto de valores M_{ij} es -1 o 0 dependiendo de si las áreas i, j se delimitan o no. Esta matriz es un objeto matrix de R con y su extensión es ".txt".

Después seleccionamos los modelos espacio temporales a usar. Podemos elegir entre los modelos espaciales iCAR, BYM o Leroux y entre los modelos temporales "paseos aleatorios" de primer o segundo orden (RW1 o RW2). Podemos además incluir o no una interacción espacio-temporal. Si no queremos interacción seleccionamos la opción Additive. En otro caso podemos elegir entre cuatro posibles interacciones (tipo Knorr-Held, Ver Ugarte et al., 2014). La interacción de tipo I no está estructurada ni en el espacio ni en el tiempo. La interacción tipo II considera estructura en el tiempo pero no en el espacio, es decir, cada región evoluciona en el tiempo según el paseo aleatorio escogido pero independientemente de sus regiones vecinas. La interacción de tipo III considera estructura en el espacio pero no el tiempo, y por último la interacción tipo IV considera estructura en el espacio y en el tiempo. Es decir, las regiones vecinas presentan evoluciones temporales similares. También podemos indicar si permitimos calcular patrones temporales, espaciales y espacio-temporales.

Finalmente elegimos la estrategia de aproximación que usará INLA para ajustar el modelo elegido. Es posible elegir tres estrategias de aproximación: Full Laplace, Simplified Laplace y Gaussian. El ajuste del modelo elegido comenzará tras pulsar el botón "Start"

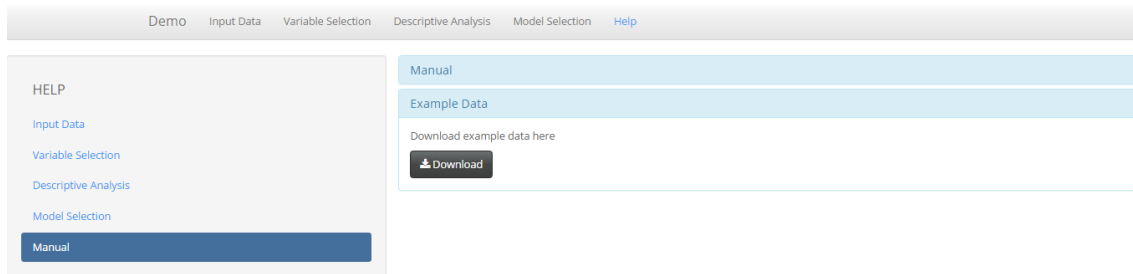
El resultado del ajuste del modelo tiene la siguiente forma:

The screenshot displays the INLA web interface. On the left, a sidebar contains configuration options: 'Select cartography' (with 'Cartography' selected), 'Spatial component' (with 'ICAR' selected), 'Temporal component' (with 'RW1' selected), 'Type of interaction' (with 'Additive' selected), and 'Strategy' (with 'gaussian' selected). A 'Start' button is at the bottom of the sidebar. The main area shows the R code for the model and its output. The output includes a table of time used, fixed effects, random effects, model hyperparameters, and various model fit statistics like DIC, WAIC, and log-likelihood.

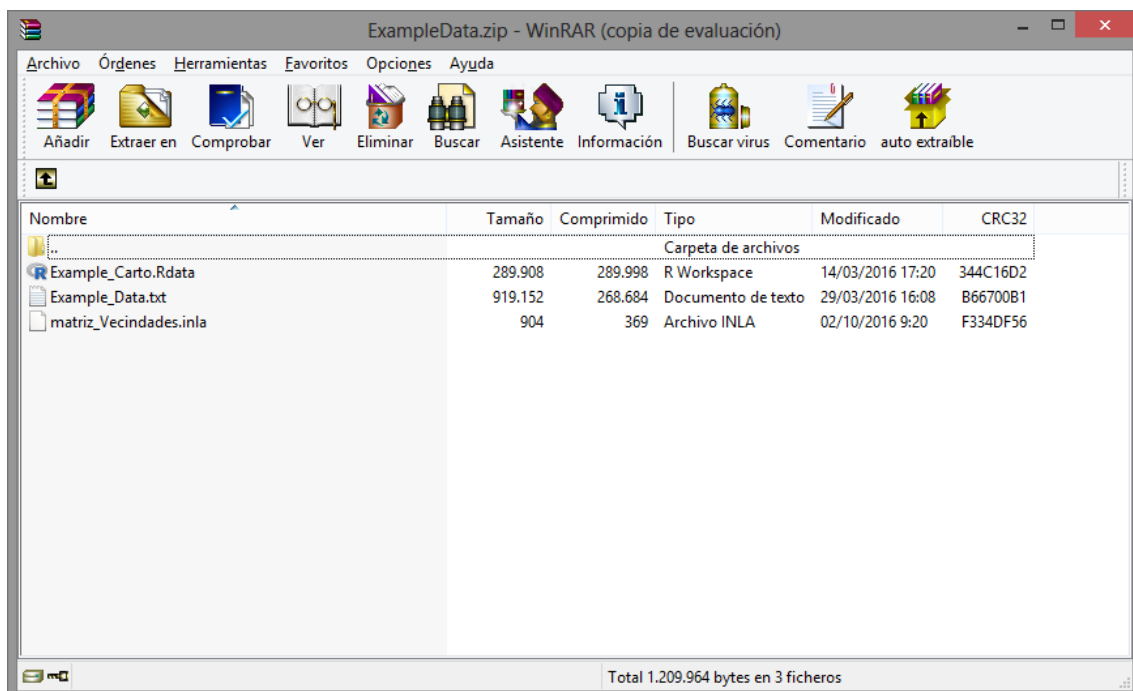
```
[1] ""
[2] "Call:"
[3] "inla(formula = formula, family = \"poisson\", data = Data.INLA, \"\", E = E, control.compute = list(dic = TRUE, cpo = TRUE, waic = TRUE), \"\",
[4] \"")
[5] "Time used:"
[6] "Pre-processing Running inla Post-processing Total "
[7] "0.2582 3.5831 0.3619 4.1992 "
[8] ""
[9] "Fixed effects:"
[10] "mean sd 0.025quant 0.5quant 0.975quant mode kld"
[11] "(Intercept) -0.0374 0.0048 -0.0469 -0.0374 -0.0278 -0.0373 0"
[12] ""
[13] "Random effects:"
[14] "Name\t model"
[15] "ID.area Generic model "
[16] "ID.year Rsd model "
[17] "ID.year1 IID model "
[18] ""
[19] "Model hyperparameters:"
[20] "mean sd 0.025quant 0.5quant 0.975quant mode"
[21] "Precision for ID.area 4.562e+01 1.079e+01 20.7794 4.380e+01 8.565e+01 38.1490"
[22] "Beta for ID.area 4.199e+01 1.070e+01 0.0903 4.021e+01 8.200e+01 0.1200"
[23] "Precision for ID.year 1.567e+03 6.835e+02 689.8759 1.445e+03 3.239e+03 1229.5600"
[24] "Precision for ID.year1 1.266e+03 7.969e+02 2552.3439 2.511e+03 8.282e+03 5529.2195"
[25] ""
[26] "Expected number of effective parameters(std dev): 59.90(1.577)"
[27] "Number of equivalent replicates : 10.48 "
[28] ""
[29] "Deviance Information Criterion (DIC) ....: 7348.41"
[30] "Effective number of parameters .....: 60.35"
[31] ""
[32] "Watanabe-Akaike Information Criterion (WAIC) ....: 7369.37"
[33] "Effective number of parameters .....: 75.87"
[34] ""
[35] "Marginal log-likelihood: -3742.02 "
[36] "CPO and PIT are computed"
[37] ""
[38] "Posterior marginals for linear predictor and fitted values computed"
[39] ""
```


Capítulo 4.6. Ejemplo de uso

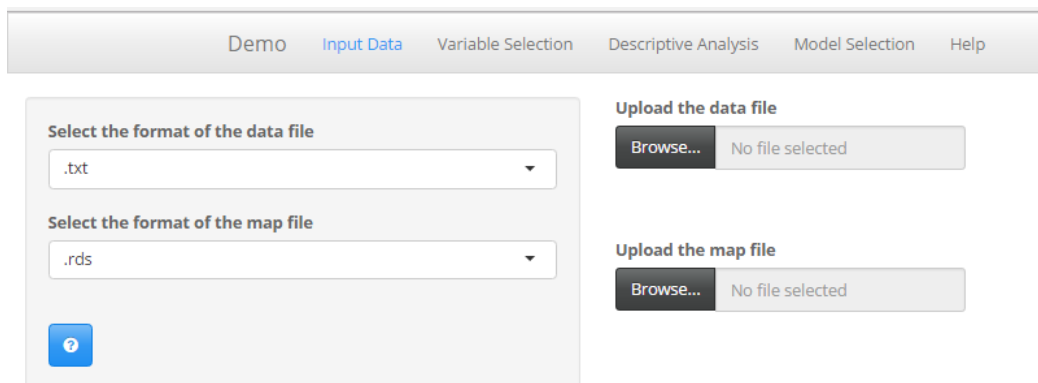
En este capítulo se realizará un ejemplo de uso de la aplicación, utilizando para ello los datos de ejemplo que pueden ser descargados en la pestaña Help.



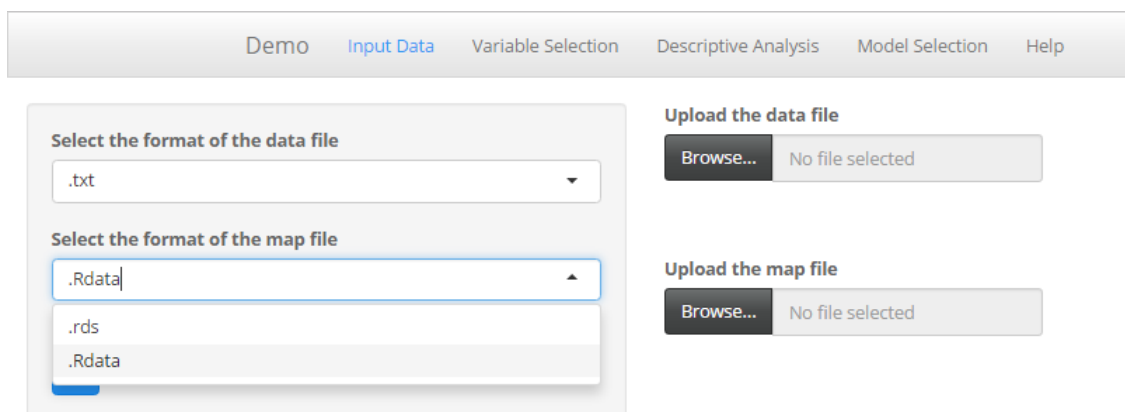
Los datos de ejemplo están contenidos en un descargable e incluyen un conjunto de datos, una cartografía de España y la matriz de vecindades asociada a esa cartografía en formato ".inla".



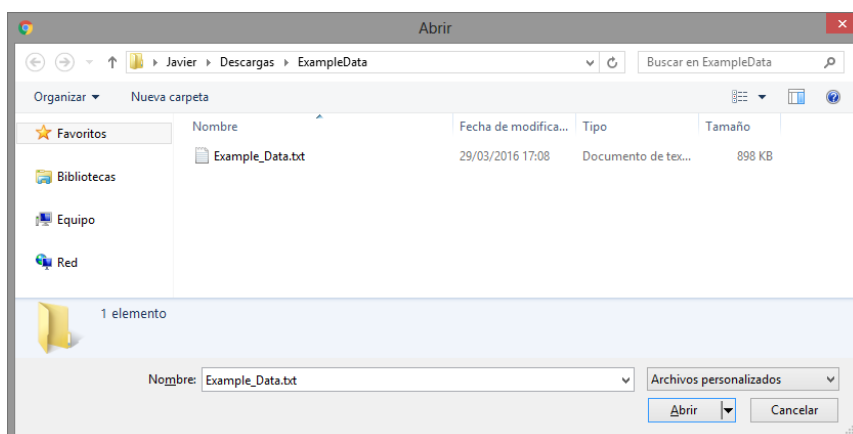
Una vez descargados los datos de prueba accedemos a la aplicación.

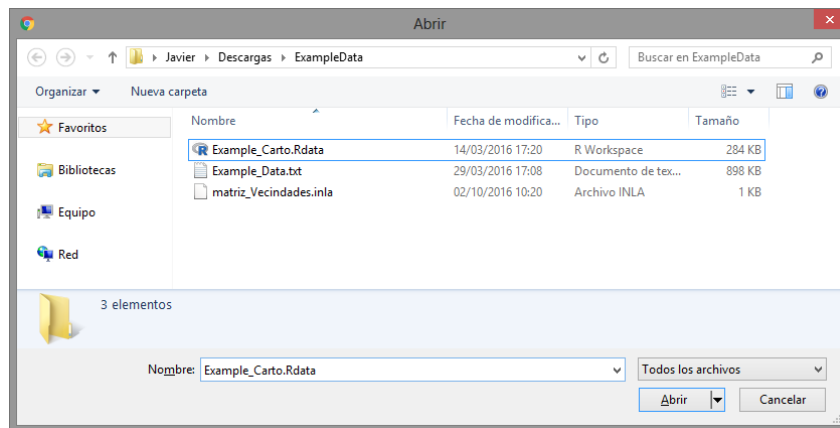


Seleccionamos los formatos de los archivos que vamos a cargar a la aplicación, ".txt" y ".Rdata".

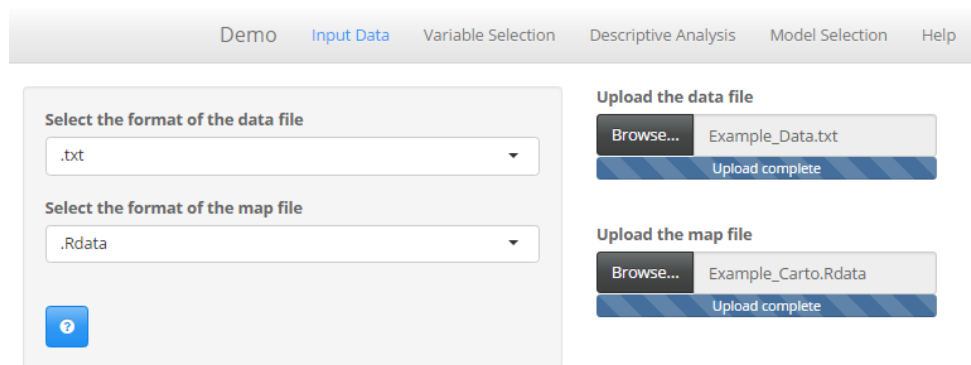


después cargamos los archivos haciendo clic en el botón "Browse" superior para cargar el archivo de datos y en el inferior para cargar el archivo con la cartografía.

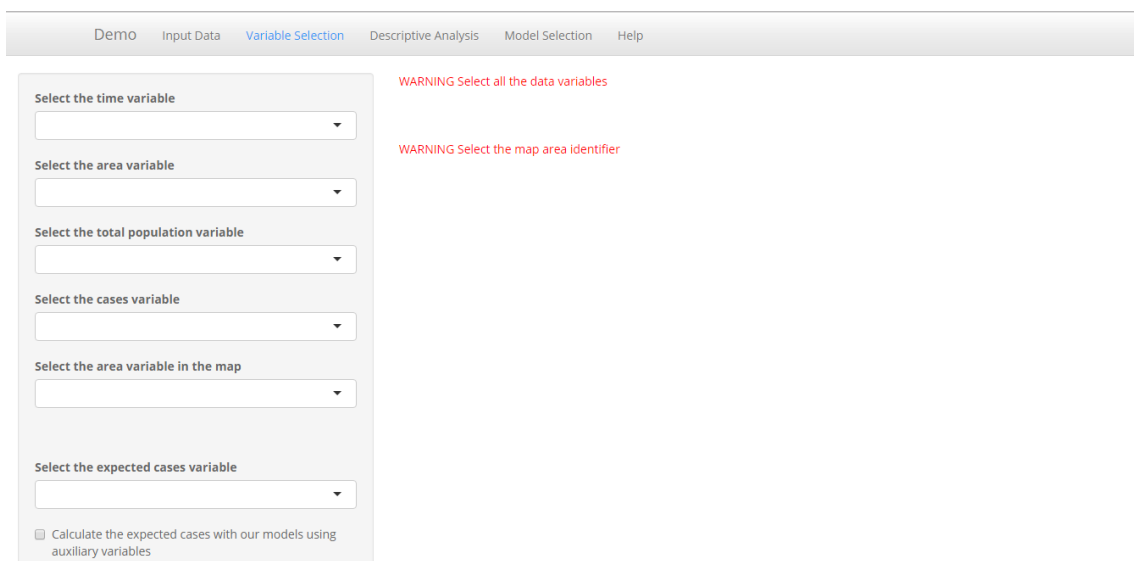




Tras cargar ambos archivos veremos lo siguiente.

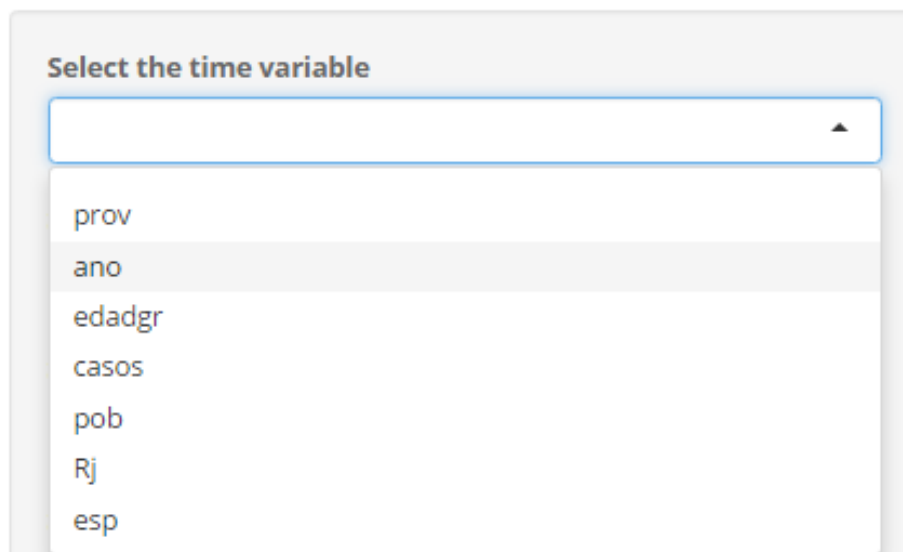


Y avanzamos al siguiente paso haciendo clic en "Variable Selection".



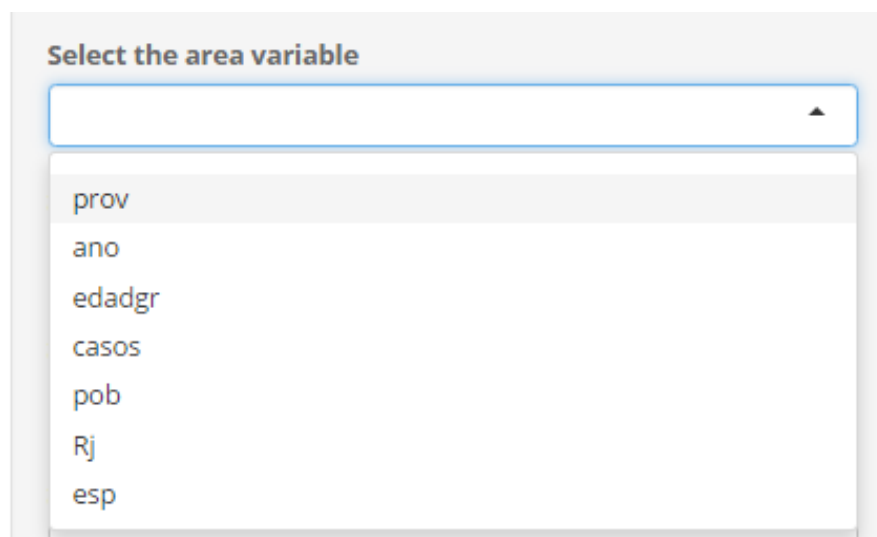
Los cuatro primeros desplegados hacen referencia a las variables del archivo de datos que representan el tiempo, el espacio (en este caso provincias), el total de la población de ese área en ese tiempo y el número de casos, respectivamente. El quinto desplegable hace referencia a las variables del archivo que contiene la cartografía. Debemos seleccionar la variable que identifica cada área. Seleccionamos las variables correspondientes de la siguiente forma:

La variable que representa el tiempo es 'ano'.



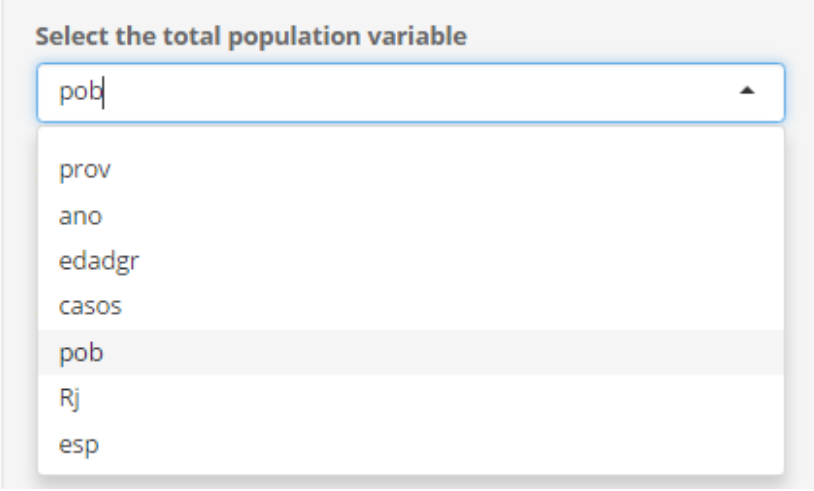
A screenshot of a web interface showing a dropdown menu titled "Select the time variable". The menu is open, displaying a list of variables: "prov", "ano", "edadgr", "casos", "pob", "Rj", and "esp". The variable "ano" is highlighted with a grey background, indicating it is the selected option.

La variable que representa el espacio es 'prov'.



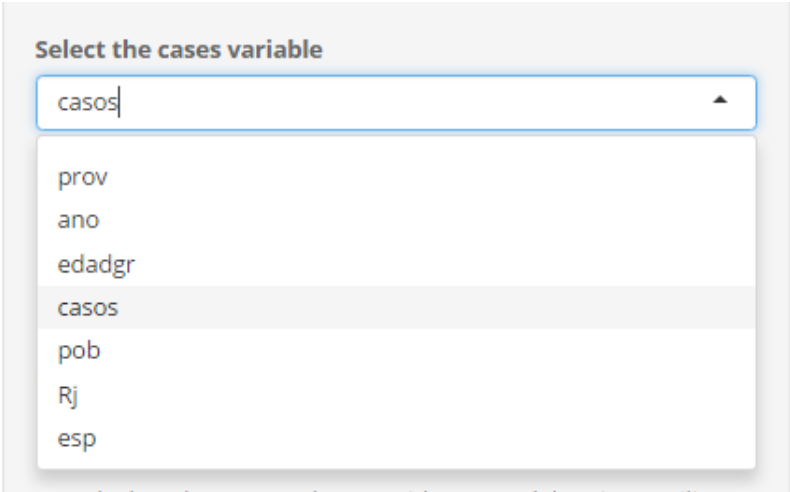
A screenshot of a web interface showing a dropdown menu titled "Select the area variable". The menu is open, displaying a list of variables: "prov", "ano", "edadgr", "casos", "pob", "Rj", and "esp". The variable "prov" is highlighted with a grey background, indicating it is the selected option.

La variable que representa la población es 'pob'.



The screenshot shows a web-based interface for selecting a variable. At the top, there is a title "Select the total population variable". Below the title is a search input field containing the text "pob". A dropdown menu is open, displaying a list of variables: "prov", "ano", "edadgr", "casos", "pob", "Rj", and "esp". The variable "pob" is highlighted with a grey background, indicating it is the selected option.

La variable que representa los casos observados es 'casos'.



The screenshot shows a web-based interface for selecting a variable. At the top, there is a title "Select the cases variable". Below the title is a search input field containing the text "casos". A dropdown menu is open, displaying a list of variables: "prov", "ano", "edadgr", "casos", "pob", "Rj", and "esp". The variable "casos" is highlighted with a grey background, indicating it is the selected option.

Tras seleccionar estas cuatro primeras variables podremos visualizar un resumen de las mismas.

Demo Input Data **Variable Selection** Descriptive Analysis Model Selection Help

Select the time variable
ano

Select the area variable
prov

Select the total population variable
pob

Select the cases variable
casos

YEAR -> Values:
1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010

AREA -> Values:
1 2 3 4 5 6 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 36 37 39 40 41 42 43 44 45 46 47 48 49 50

POPULATION -> Range:
1496 286088

CASES -> Range:
0 168

Error: undefined columns selected

A continuación seleccionamos el identificador de área de la cartografía.

Select the area variable in the map

ID.area

ID.area

NAME

Si los identificadores de las áreas no coinciden entre datos y cartografía recibiremos un aviso por parte de la aplicación, ya que los identificadores deben coincidir para que la aplicación funcione correctamente.

Ahora debemos elegir la variable que contiene los casos esperados, en caso de que ya estén calculados o indicar qué variables auxiliares queremos que se usen para calcularlos. Para este ejemplo se calcularán los casos esperados considerando como variable auxiliar los grupos de edad 'edadgr'.

☒ Calculate the expected cases with our models using auxiliary variables

Add new auxiliary variable

Delete auxiliary variable

Variable number 1

edadgr

prov

ano

edadgr

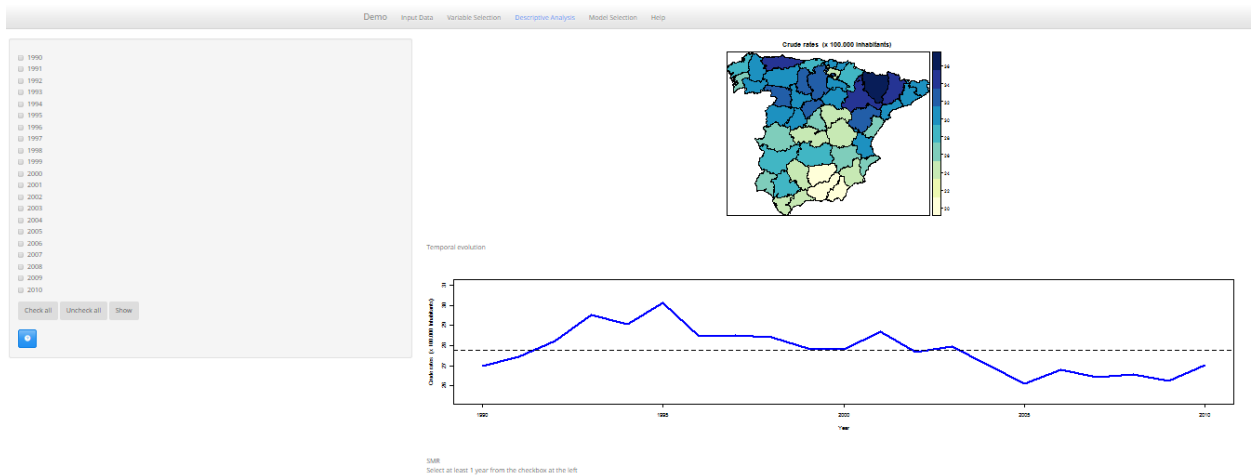
casos

pob

Rj

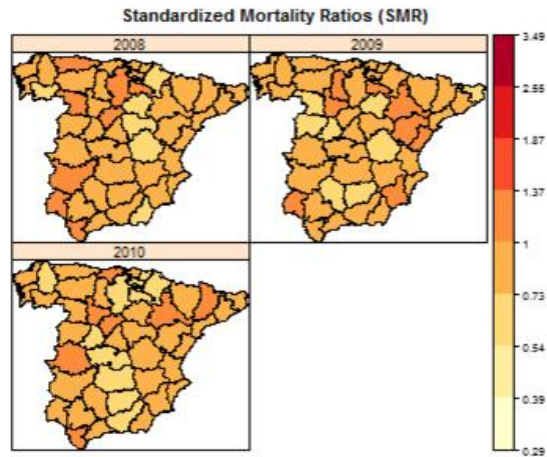
esp

Y avanzamos al siguiente paso haciendo clic en "Descriptive Analysis". Tras unos segundos veremos la representación de las tasas brutas.



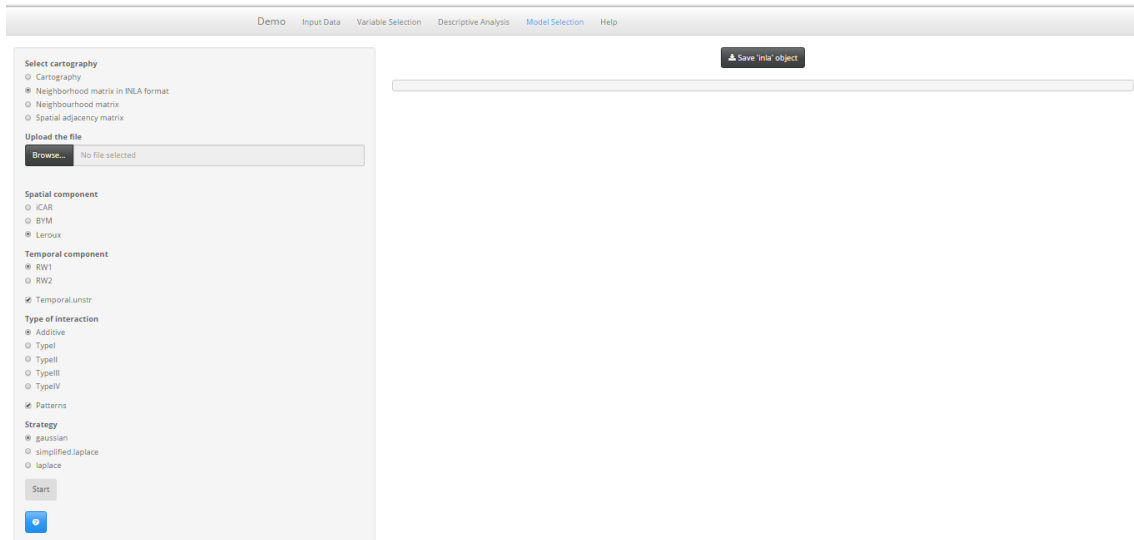
Para este ejemplo calcularemos los SMR de los 3 últimos años.

SMR

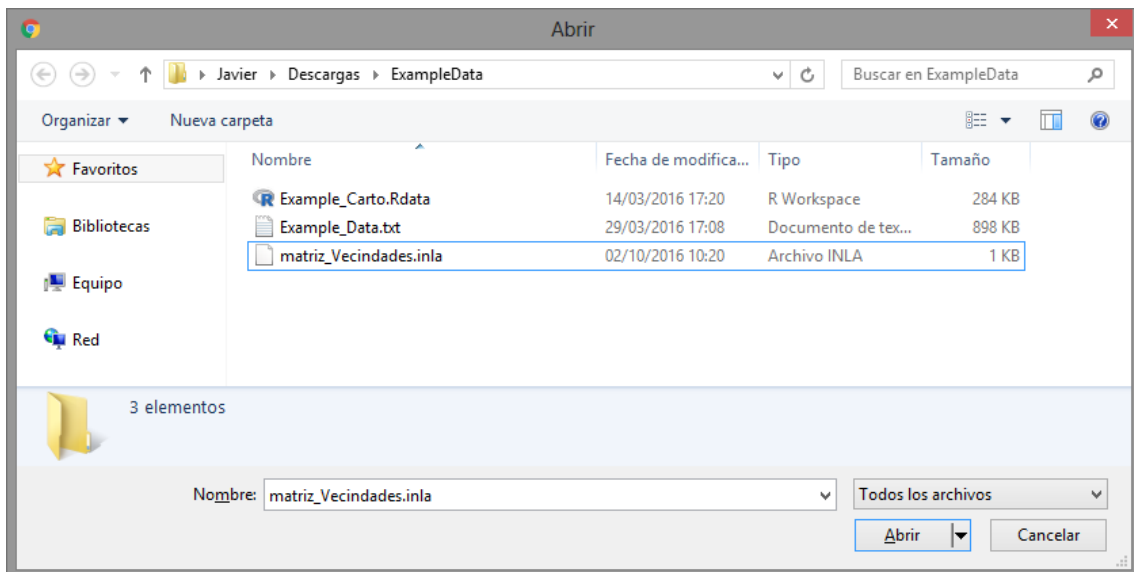


[Download](#)

Y avanzamos al siguiente paso haciendo clic en "Model Selection".



Para este ejemplo usaremos la matriz de estructura espacial incluida en los datos descargados al empezar el ejemplo.



Para obtener el mismo modelo que en este ejemplo seleccione las siguientes opciones.

Select cartography

- ☐ Cartography
- ☒ Neighborhood matrix in INLA format
- ☐ Neighbourhood matrix
- ☐ Spatial adjacency matrix

Upload the file

matriz_Vecindades.inla

Upload complete

Spatial component

- ☐ iCAR
- ☐ BYM
- ☒ Leroux

Temporal component

- ☐ RW1
- ☒ RW2
- ☒ Temporal.unstr

Type of interaction

- ☐ Additive
- ☐ TypeI
- ☐ TypeII
- ☐ TypeIII
- ☒ TypeIV
- ☒ Patterns

Strategy

- ☐ gaussian
- ☒ simplified.laplace
- ☐ laplace

Y hacemos clic en "Start", en unos minutos obtendremos el siguiente resultado que podemos descargar haciendo clic en "Save 'inla' object".

```

INLA version .....: 0.0-1468872408
INLA date .....: Mon 18 Jul 22:06:48 CEST 2016
INLA hgid .....: hgid: 895e86f8a94e date: Mon Jul 18 14:43:05 2016 +0100
INLA-program hgid .....: hgid: 895e86f8a94e date: Mon Jul 18 14:43:05 2016 +0100
Web-page .....: www.r-inla.org
Email support .....: help@r-inla.org
                  : r-inla-discussion-group@googlegroups.com
Source-code .....: bitbucket.org/hrue/r-inla

Time used:
Pre-processing   Running inla Post-processing   Total
      0.3082         207.9618           0.8386      209.1085

Fixed effects:
      mean      sd 0.025quant 0.5quant 0.975quant   mode kld
(Intercept) -0.0383 0.0051   -0.0483   -0.0382   -0.0283 -0.0382  0

Random effects:
Name      Model
ID.area   Generic1 model
ID.year   RiI2 model
ID.year1   IID model
ID.area.year   Generic0 model

Model hyperparameters:
      mean      sd 0.025quant 0.5quant 0.975quant   mode
Precision for ID.area   4.517e+01 1.535e+01 2.202e+01 4.201e+01 8.186e+01 3.871e+01
Beta for ID.area        4.052e-01 1.725e-01 1.174e-01 3.914e-01 7.606e-01 3.416e-01
Precision for ID.year   1.052e+04 1.074e+04 1.240e+03 7.372e+03 3.870e+04 3.345e+03
Precision for ID.year1   1.306e+04 1.697e+04 1.393e+03 7.980e+03 5.584e+04 3.465e+03
Precision for ID.area.year 7.037e+04 5.806e+04 1.436e+04 5.404e+04 2.239e+05 3.332e+04

Expected number of effective parameters(std dev): 115.16(5.74)
Number of equivalent replicates : 8.57

Deviance Information Criterion (DIC) ....: 7228.02
Effective number of parameters .....: 116.80

Watanabe-Akaike information criterion (WAIC) ....: 7235.52
Effective number of parameters .....: 111.08

Marginal log-Likelihood: -4508.15
CPO and PIT are computed

Posterior marginals for linear predictor and fitted values computed

```


Anexos

Una versión de esta aplicación se presentó en las VIII Jornadas de Usuarios de R.



CERTIFICADO DE PRESENTACIÓN ORAL

Albacete, 22 de noviembre de 2016

Este certificado es para confirmar que el trabajo titulado **Una aplicación web interactiva en Shiny para el análisis espacio-temporal de riesgos de mortalidad en áreas pequeñas** de los autores *A. Adán, J.M. Carrillo, M.D. Ugarte* se ha presentado como charla en las VIII Jornadas de Usuarios de R celebradas los días 17 y 18 de noviembre de 2016 en la Agrupación Politécnica del campus de Albacete de la Universidad de Castilla-La Mancha.

Virgilio Gómez-Rubio, UCLM
Coordinador del Comité Organizador Local

<http://www.r-es.org/8jornadasR>

Organizadoras



Patrocinadores por



Bibliografía

- [1] Allaire, J.J., Cheng, J., Xie, Y., McPherson, J., Chang, W., Allen, J., Wickham, H., Allaire, J. and Hyndman, R. (2016). rmarkdown: Dynamic Documents for R. R package version 1.0.
<https://CRAN.R-project.org/package=rmarkdown>
- [2] AnalytixWare (2014). shinySky:
A set of Shiny UI components/widets R package version 0.1.2.
<https://github.com/AnalytixWare/ShinySky>
- [3] Bailey, E. (2015). shinyBS: Twitter Bootstrap Components for Shiny. R package version 0.61.
<https://CRAN.R-project.org/package=shinyBS>
<https://ebailey78.github.io/shinyBS/>
- [4] Bivand, R.S., and Lewin-Koh, N. (2016). maptools: Tools for Reading and Handling Spatial Objects. R package version 0.8-39.
<https://CRAN.R-project.org/package=maptools>
- [5] Bivand, R.S., Pebesma, E., Gomez-Rubio, V. (2013). *Applied spatial data analysis with R*, Second edition. Springer, NY.
<http://www.asdar-book.org/>
- [6] Bivand, R.S., Hauke, J., and Kossowski, T. (2013). Computing the Jacobian in Gaussian spatial autoregressive models: An illustrated comparison of available methods. *Geographical Analysis*, **45(2)**, 150-179.
- [7] Bivand, R.S., Piras, G. (2015). Comparing Implementations of Estimation Methods for Spatial Econometrics. *Journal of Statistical Software*, **63(18)**, 1-36.
URL <http://www.jstatsoft.org/v63/i18/>.

- [8] Chang, W. (2016). shinythemes: Themes for Shiny. R package version 1.1.1.
<https://CRAN.R-project.org/package=shinythemes> de RStudio
- [9] Chang, W., Cheng, J., Allaire, J.J., Xie, Y. and McPherson, J. (2016). shiny: Web Application Framework for R. R package version 0.14.2.
<https://CRAN.R-project.org/package=shiny>
- [10] Lindgren, F., Rue, H. (2015). Bayesian Spatial Modelling with R-INLA. *Journal of Statistical Software*, **63 (19)**, 1-25.
URL <http://www.jstatsoft.org/v63/i19/>.
- [11] Lindgren, F., Rue, H., and Lindstrom, J. (2011). An Explicit Link Between Gaussian Fields and Gaussian Markov Random Fields: The Stochastic Partial Differential Equation Approach (with discussion). *Journal of the Royal Statistical Society B*, **73 (4)**, 423-498.
- [12] Martins, T.G., Simpson, D., Lindgren, F. and Rue, H. (2013). Bayesian computing with INLA: New features. *Computational Statistics and Data Analysis*, **67 (2013)** 68-83.
- [13] Neuwirth, E. (2014). RColorBrewer: ColorBrewer Palettes. R package version 1.1-2.
<https://CRAN.R-project.org/package=RColorBrewer>
- [14] Pebesma, E.J., Bivand, R.S., 2005. Classes and methods for spatial data in R. *R News* 5 (2).
<http://cran.r-project.org/doc/Rnews/>.
- [15] Repositorio oficial de paquetes de R
<https://cran.r-project.org/>
- [16] Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian Inference for Latent Gaussian Models Using Integrated Nested Laplace Approximations (with discussion). *Journal of the Royal Statistical Society B*, **71**, 319-392.

- [17] Ugarte, M.D., Adin, A., Goicoa, T., Militino, A.F. (2014). On fitting spatio-temporal disease mapping models using approximate Bayesian inference. *Statistical Methods in Medical Research*, **23**:507 - 530.

- [18] Web oficial de BootStrap
<http://getbootstrap.com/>

- [19] Web oficial del proyecto R-INLA
<http://www.r-inla.org/>

- [20] Web oficial de R
R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
URL <https://www.R-project.org/>.

- [21] Web oficial de StackOverflow
<http://stackoverflow.com/>

- [22] Webs oficiales de RStudio
<https://www.rstudio.com/>
<https://blog.rstudio.org/>
<http://shiny.rstudio.com/>
<http://www.shinyapps.io/>